

РОЗРОБКА ІНФОРМАЦІЙНОЇ СИСТЕМИ АВТОМАТИЗОВАНОЇ ЛЕМАТИЗАЦІЇ ДІЄСЛІВ НІМЕЦЬКОЇ МОВИ

Борисова Н.В., Оліфенко І.В.

*Національний технічний університет
«Харківський політехнічний інститут», м. Харків*

Розробка системи автоматизованої лематизації є актуальною тому, що процедура лематизації використовується для вирішення багатьох задач обробки природної мови. Лематизація – це один з методів морфологічного аналізу, метою якого є приведення словоформи до її основної словникової форми (леми). У результаті лематизації від словоформи відокремлюються флексивні закінчення та утворюється основна форма слова (наприклад, для іменників – це називний відмінок, однина; для прикметників – називний відмінок, однина, чоловічий рід; для дієслів – інфінітивна форма).

У роботі розглянуто інформаційну систему автоматизованої лематизації дієслів з відокремлюваними префіксами у німецькій мові. Складність розробки та роботи такої системи полягає, по-перше, у тому, що відокремлюваний префікс може знаходитися на відстані у декілька слів від дієслова, до якого він належить, у кінці речення або у кінці простого речення у складі складнопідрядного, а, по-друге, у тому, що відокремлюваний префікс може частково або повністю змінювати значення дієслова, до якого він приєднується, доповнювати, уточнювати його і т.п. Також при створенні інформаційної системи, крім особливостей дієслів з відокремлюваними префіксами, необхідно зважати на загальні правила та особливості формотворення дієслів у німецькій мові, адже від їх врахування залежить правильність самої лематизації.

У якості математичного забезпечення інформаційної системи використано модель формотворення природної мови, запропоновану Пруцковим О.В. [1]. Вона розглядає генерацію будь-якої словоформи з певним граматичним значенням як послідовність кінцевого числа прямих або зворотних перетворень основи. Для вирішення задачі лематизації доцільно застосовувати до словоформ зворотне перетворення. На базі цієї моделі побудовано алгоритм процесу автоматизованої лематизації німецьких дієслів з відокремлюваними префіксами, також до алгоритмічного забезпечення інформаційної системи входить алгоритм знаходження дієслів з відокремлюваними префіксами у реченнях.

Для створення інформаційної системи використано мову програмування Java. Для здійснення перекладу дієслів підключено онлайн-перекладач Google та електронний німецько-російський словник.

Література:

1. Пруцков А. В. Модели, методы и программы автоматической обработки форм слов в естественно-языковых интерфейсах: дис. на соискание учен. степ. доктора техн. наук: спец. 05.13.11 – математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей / А. В. Пруцков. – Рязань: Рязанский государственный радиотехнический университет, 2015. – 253 с.